DOCUMENT RESUME

ED 455 281 TM 033 084

AUTHOR Arce-Ferrer, Alvaro J.; Cisneros-Cohernour, Edith J.
TITLE Standardized Assessment in Mexico: Issues on Interpretation and Use of Assessment Results.
PUB DATE 2001-04-00
NOTE 25p.; Paper presented at the Annual Meeting of the American Educational Research Association (Seattle, WA, April 10-14, 2001).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Culture Fair Tests; English; Foreign Countries; *High School Students; High Schools; *Measures (Individuals); Psychometrics; Reliability; *Self Efficacy; *Spanish; Test Construction; *Translation; Validity
IDENTIFIERS Career Decision Making Self Efficacy Scale; Emic Etic Analysis; *Mexico

ABSTRACT

This paper summarizes main findings from a two-step investigation of the translation of a psychological scale from English into Spanish. The overall purpose of the study was to document the effects of tailoring a scale with etic items (i.e., culturally general items) and emic items (i.e., culture specific items) on the quality of the information. Other goals were to document the psychometric properties of self-efficacy scores for a sample of Spanish-speaking students from southern Mexico and to appraise the effects of extreme response style on total, factor, and item congruence. The short form of the Career Decision Making Self-Efficacy Scale (CDMSE-SF) (Betz, Klein, and Taylor, 1996) was tailored and administered to 3,000 high school students. Reliability and validity evidence was gathered using standard psychometric practices. For the second part of the study, the scale was administered to two samples of students differing in their extreme response styles. Total, factor, and item congruence coefficients were gathered from Procrustes rotation. Reliability and validity evidence did not support the five-dimension structure of the self-efficacy construct. Reliability estimates for the translation were markedly inferior to those reported for the source language, and dimensions present in the source culture were not reproducible in the target culture. Students could not distinguish between items dealing with self-appraisal and items dealing with goal selection. The second part of the study corroborated these findings. Nonuniform effects of the extreme response style were found on etic and emic items. The observed no invariance might relate more to extreme response styles than self-efficacy. (Contains 8 tables and 34 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 455 281

RUNNING HEAD: Translation of psychological scales, linguistic similarity, psychometric equivalence across groups, test score interpretation

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
E. Cisneros - Cohernour
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
1

Standardized Assessment in Mexico: Issues on interpretation and use of assessment results

Alvaro J. Arce-Ferrer
Universidad Autonoma de Yucatan

and

Edith J. Cisneros-Cohernour
Universidad Autonoma de Yucatan
University of Illinois at Urbana-Champaign

TM033084

Copies of this paper are
available from:
Alvaro J. Arce-Ferrer
Universidad Autonoma de Yucatan, Mexico
aferrer@tunku.uady.mx

Paper presented at the Annual Meeting of the American Educational Research Association (División D). Seattle, Washington, April 11, 2001.

BEST COPY AVAILABLE

ABSTRACT

This paper summarizes main findings from a two-step investigation about translation of a psychological scale from English into Spanish. The overall purpose of the research was to document the effects of tailoring a scale with etic items (i.e., culturally general items) and emic items (i.e., culture specific items) on the quality of the information. Particularly, the goals were to document the psychometric properties of self-efficacy scores for a sample of Spanish-speaking students from the southern region of Mexico and to appraise the effects of extreme response style on total, factor, and item congruence.

The Carrer-Decision Making Self-Efficacy scale was chosen because it comprises items dealing with relevant behaviors in both languages (source for etic items). Additionally, it oversees some practices relevant in the target culture (source for emic items). Suggestions from Hambleton (1994), Allalouf, Hambleton and Sireci (1999), and Hulin (1987) guided the process for item selection and development for etic and emic items.
The tailored scale was administered to a sample of 3000 senior high school students and reliability and validity evidences were gathered using standard psychometric practices. For the second part, the scale was administered to two samples of students differing in their amounts in extreme response styles estimates. Total, factor, and item congruence coefficients were gathered from procrustes rotation (Cliff, 1966; Schönemann, 1966; McCrae et al., 1996).

Both reliability and validity evidence did not support the five dimension structure of the self-efficacy construct. Reliability estimates were marginally inferior compared to those reported for the source language. Dimensions present in the source culture were not reproducible in the target culture. Students could not distinguish between items dealing with self-appraisal and items dealing with goal selection. The second part of the study corroborates the preceding findings.
Nonuniform effects of the extreme response style were found on etic and emic items. The observed noninvariance might relate more to extreme response styles than self-efficacy.

INTRODUCTION

High fidelity in translated testing is more than achieving linguistic similarity between source and target languages. It has been fully documented the limited equivalence of translated psychological scales (Angoff and Cook, 1988; Hulin, 1987; Allalouf, Hambleton, and Sireci, 1999). All the gathered evidence indicates that linguistic translation is a necessary, but not sufficient condition to ensure equivalence between source and target languages. In this regard, current Standards for Educational and Psychological Testing (AERA, APA, and NCME, 1999) states:

> "... one cannot simply assume that such translation produces a version of the test that is equivalent in content, difficulty level, reliability, and validity to the original untranslated version. Further, one cannot assume that test takers' relevant acculturation experiences are comparable across the two versions."

To keep fidelity high, translation should render similar linguistic and measurement characteristics of items and scales. The goal of test adaptation is to ensure invariance of the construct and its representation across languages (Hambleton, 1994). Construct invariance of psychological scales across countries is a validation of the theory upon which the scale relies. Additionally to threats from undetected errors in translations and coding blunders, for example, score interpretation faces several challenges when generalizing a construct from one culture to another. First, construct definition needs to be assessed before engaging in an investigation of construct and factorial invariance. A well defined construct in culture A might be ill defined in culture B. For example, a measure of job satisfaction developed in culture A, which values worker independence, might not include items tapping compliance to supervisors, valued as important in culture B. Second, balance of the construct is another aspect to investigate in cross-cultural test adaptation studies. Construct balance might be lost when test translation overlooks some important cultural characteristics. For instance, in a measure of competency for testing, the number of items assessing knowledge and skills of testing accommodation practices might seem too many in a culture less committed to diversity.

Third, sources of construct contamination need to be flagged, also. Construct contamination happens when an attribute other than the one measured becomes an active part of the measurement process. Sources of construct contamination have typically arisen from word and sentence difficulty, change in content and format, and cultural irrelevance (Allalouf, Hambleton, and Sireci, 1999). In cross-country research, other sources of construct contamination might arise from differences in respondents' information processing (Cheung and Rensvold, 2000; Little, 2000). For some cultures, styles of decoding, encoding, and answering items might be different from other cultures.

Two sources for construct contamination are perceived "tone" in the statement and response style. The former triggers emotional answers from members of a society, which assess tone as part of information decoding. Tone in a statement might sound polite in one culture, but rude in another. The latter narrows a score scale. For example, people wishing to demonstrate sincerity and conviction would score closer to the upper or lower part of a scale than cultures in which modesty and nonjudgment are valued. The preceding response pattern is known as extreme response style (Clarke, 2000; Gibbons, Zellner, and Rudek, 1999; Mullen, 1995).

As part of the psychometric literature, there are strategies to deal with some of the above threats to validity in cross-cultural research. Most of them depart from uncritically taking mean differences between scales administered to members from source and target cultures (Arce-Ferrer, 2000; Cheung and Rensvold, 2000). Some approaches are based on judgmental evidence and others are based on empirical evidence. Judgmental sources of evidence render valuable information about possible explanations for translated items function differentially (Hambleton and Jones, 1995). Methods based on empirical evidence are valuable for flagging errors of translation and cultural inadequacies in translated items (Hulin, 1987, Ellis, 1995).

The use of tailored tests has been advocated for increasing cultural relevance, also. For example, Hulin (1987) mentions some advantages of tailored tests from well-translated etic items and new emic items. Etic items are those found appropriate across cultures; whereas the emic items are culture specific. A tailored test might experience the effects of sources of construct contamination such as differences in respondents' information processing styles. Some studies have found tailored tests containing both emic and etic items only marginally superior compared to those containing only etic items (deVera, 1985). There is little research exploring why a tailored test, which supposedly has the strengths from both cultures, performs only slightly better. Few studies have tried to deal with the relationship between differences in respondents' information processing and performance of tailored tests. A current revision of these studies can be found in Cheun and Rensvold (2000).

PURPOSE

The primary purpose of this paper is to study the quality of the information collected from a scale tailored with etic items (linguistically and culturally adapted items) and emic items (newly developed items directly in the target language) for Spanish speaking people. The main objectives are to:

1. Describe the validity and reliability evidence of a psychological scale tailored from etic and emic items.

2. Evaluate effects of extreme response style on total, factor, and item congruency.

3. Evaluate whether the effects of extreme response style on emic and etic items are uniform (affecting all responses) or nonuniform (affecting some responses.)

METHODOLOGY

Test selection

The short form of the Career Decision-Making Self-Efficacy Scale (CDMSE-SF; Betz, Klein, and Taylor, 1996) was examined. The CDMSE-SF is a 25 item scale asking students to rate their confidence in their ability to perform each of the career decision-making tasks on a scale of 0 (no confidence) to 9 (complete confidence.)

The CDMSE-SF test was chosen for several reasons. First, career-decision making is a construct, which explores competencies highly relevant in the source culture. To measure the career self-efficacy construct, the CDMSE-SF samples significant behaviors culturally appropriate for the US. Second, the psychometric properties of the form of the CMDSE-SF scale have been established. The scale utilizes samples of behaviors from a theoretical model proposed to explain career maturity (Crites, 1978). The instrument explores five career choice competencies: (a) accurate self-appraisal, (b) gathering occupational information, (c) goal selection, (d) making plans for the future, and (e) problem solving. Validity and reliability evidences were found comparable or better than those in the long form of the test (Betz, Klein & Taylor, 1996). Third, validity and reliability evidences of the CDMSE-SF, as well as for its long form, have been documented for different genders, ethnic background, and vocational certainty (Betz & Luzzo, 1996; Fouad, Smith, Enochs, 1997; Luzzo, 1993). However, there is no validity and reliability evidence established for cultures like the Mexican one. Fourth, the Career Decision-Making Self-Efficacy has appealing characteristics for vocational orientation in Mexico (Canto, 1999; Ramirez, Arce-Ferrer & Canto, 2000). However, there is no culturally developed instrument, which could be used.

Context and Subjects

Different from a typical US high school program, Mexico's senior high school programs are seen as pre-college training programs. Students at this level have completed six years of elementary school and three years of junior high school. Training at this level last for three year. The first two years are devoted to developing general academic skills and acquiring general knowledge. The last year is devoted to introducing discipline specific knowledge in one of the following four areas: (1) biology, (2) mathematics, (3) business and administration, or (4) social studies and humanities. At the end of the third year, students apply for college admission in a career path congruent with the chosen discipline in the last year of their senior high school program. For example, students majoring in mathematics can apply for admission to those colleges requiring training in mathematics, but not for careers requiring majors in biology or business and administration, for example (Canto, 1999).

The participants in the study were male and female senior high school students from the southern region of Mexico. Two different samples were used in this study. For the first part of the study, a total sample of 3,153 high school students participated in the study. Approximately one half of the total sample (n=1501) were males and the remaining (n=1652) students were females. A little more than one third of the overall sample was enrolled in the first year of senior high school (n=1,299) and the remaining were enrolled in the second and third year, respectively (n=956, n=989).

For the second part of the study, documenting the invariance across two groups, an independent sample of 400 high school students was obtained. Special consideration was followed to match relevant variables present in the first sample, such as gender and age. The sample contained two hundred students were certain about their major an equal number of students not yet certain about their major. The rationale for choosing these two groups was to introduce the extreme response style variable. Mexican culture expects for those having made a decision to demonstrate sincerity and conviction in spite of future changes (Díaz-Guerrero, 1975). Table 1 summarizes descriptive information of the second sample.

Insert Table 1 here

Tailoring the test

Choosing etic items

The translation process focused on obtaining linguistic similarity between the source and the target language. The flow diagram by Allaouf, Hambleton and Sireci (1999) guided the process for translating items from the source to the target language. To decrease the cognitive requirements for translators, it was decided to implement the procedure in two separated stages. In the first stage, the purpose was to obtain linguistic similarity between the translated version and the source version of the CDMSE-SF. The back-translation approach (Brislin, 1970) was utilized and issues covered in this stage were related with correctness of the translation, format equivalence, and level of difficulty on the chosen words. As part of the second stage, cultural relevance of item content was taken into account as part of the translation. Issues covered at this stage relate with relevance of behaviors to senior high school students from southern Mexico.

The steps to achieve linguistic similarity in the translation of the CDMSE-SF included the following:

1. Three bilingual instructors from a program for teaching English as a second language in a southeast public university translated the test from the source to the target language. Instructions given to translators focused on attaining linguistic similarity. They were asked to conduct a literal translation and when no direct word was available into the target language, they were asked to deviate from direct translation. These deviations were documented for future research.

2. Independent translations were then compared to reach consensus. For each translated item, faculty discussed their translation. Consensus was reached after debating about linguistic resemblance between target and source item versions.

3. Another sample of three independent instructors was asked to translate back to the source language. Similar instructions to achieve linguistic similarity as those given to the previous sample of translators were given to this current sample of translators.

4. Discrepancies in translation were resolved by agreement among the independent translators.

5. Three vocational orientation counselors appraised cultural relevance of behaviors for each of the 25 items in the target language. There were unique aspects of the US vocational orientation process that could not be kept as part of the scale. For example, one of the source items content dealt with confidence in preparing a resume. This skill, even though important when finding a job, was judged irrelevant when choosing an area of concentration for the last year of senior high school. Counselors maintained or modified the translated items based on current vocational practices in Mexico. From this analysis, 17 out of 25 items in the scale were found culturally relevant. To

supplement this set of items, another set with 8 items was chosen from the long version of the CDMSE scale using a similar approach.

6. The vocational orientation experts worked to increase the relevance of the set of 25 chosen items. It was observed that some of the behaviors were culturally relevant after making item content more general. For example, a source item dealing with finding information in the library about occupations you are interested in was assessed as containing uncommon behavior for Mexico's context. Occupational information is unlikely to be part of the information available in most public libraries. Consequently, the item was made relevant by phrasing it more generally, "find information about occupations you are interested in."

Developing emic items

It was judged convenient to create new items directly from the target language to meet the cultural relevance criterion. This strategy will allow supplemental information gathered from those items adapted by making them more general. There are unique aspects of Mexico's vocational orientation process not sampled by the etic item set. For example, it is common to find vocational activities hosting guest speakers who verbally present career information such as challenges, requirements, job opportunities, etc. School counselors were invited to write a sample of new items covering unique aspects of Mexico's vocational orientation practices. Table 2 displays information about number of etic and emic items present in the tailored scale.

Insert Table 2 here

Factor congruency

The orthogonal procrustes rotation method was followed to investigate factor congruency. Procrustes transformation is the simplest form of confirmatory factor analysis (Korth and Tucker, 1976). This transformation has been found to have the edge over other approaches, such as confirmatory factor analysis, when studying factor invariance across different populations (McCrae, et all, 1996; Paunonen, 1997). It does not require specifications of a simple structure comprised by high item loadings for factor indicators and null loadings for the other indicators as it is usually required by confirmatory studies (Rensvold and Cheung, 1998; McCrae, et all, 1996). However, among its limitations, procrustes transformation has been criticized by lacking a known sampling distribution to support hypothesis testing (Chan, Ho, Leung, Chan, and Yung, 1999). The literature has furnished some rules of thumb to overcome such a limitation (Paunonen, 1997). McCrae's SAS/IML codes were modified to conduct the procrustes rotation and to estimate the congruency coefficients for the total structure, factors, and items.

RESULTS

The tailored scale and the translated version of the CDMSE-SF were spiraled to a random sample of students from the population of this study. The mean testing time was 15 minutes, approximately. Table 3 displays reliability information for source and target languages.

Insert Table 3 here

From Table 3 it can be observed that the tailored scale performed slightly below the long version of the CDMSE scale in the source language. Reliability estimates for total scale differs by 0.02. This difference is of negligible significance. At the subscale level, tailored and source versions displayed similar performance in four out of five subscales. For the subscale dealing with accuracy in self-appraisal, reliability estimates for the tailored scale was considerably below its counterpart.

Other relevant information in Table 3 indicates the psychometric superiority of the CDMSE-SF in the source language to its linguistically similar counterpart. From Table 3, markedly differences in sizes of reliability estimates between CDMSE-SF for source and target languages can be observed. Whereas the reliability for the total scale was 0.94 for the source language, the scale reliability after translation to achieve linguistic similarity diminished to 0.88. It is observed that the largest discrepancies between the two versions arose from goal selection and problem solving subscales.

Finally, Table 3 allows making comparisons between the CDMSE-SF in its source language with the tailored version. It was striking to find the minor effect of almost doubling the number of items on the total scale reliability estimate. Regarding the total scale reliability, both versions of the test achieved almost the same estimated value; even though the former version had less than half of the items than the latter version. However, in some subscales, such as problem solving and gathering occupational information, the tailored version yielded larger estimates of reliability which might be caused by just simply augmenting the number of items in such scales or by using culturally appropriate items or by both.

Construct validity

An exploratory factor analysis of the tailored scale was performed to investigate (1) to what extent the general factor structure of the self-efficacy construct might be, (2) to document how the empirically derived factors correspond to the five dimensions which the test ostensibly measures, and (3) to document the performance of etic and emic items. The correlation matrix with squared multiple correlations in its main diagonal as initial estimates of commonalties was factor analyzed using the principal components analysis (Merenda, 1997). Based on the original design of the CDMSE scale, five factors were extracted, one for each of the five areas of competencies. The five factor solution accounted for 42.2% of the total variance and each of the five factors explained 30.4, 3.9, 2.7, 2.8, and 2.4, respectively. Given the low variance accounted for by the model, a varimax rotation was performed. Table 4 summarizes results from rotating the principal component factorial solution.

Insert Table 4 here

Table 4 displays factor loading for the fifty items comprising the scale. After rotating the factor structure, the presence of five dimensions was noticeable. The first dimension comprises items dealing with both accurate self-appraisal and goal selection subscales. Out of 20 items, 12 showed factor loading estimates greater than 0.40. For interpretational clarity, a salient loading of 0.40 was selected as one that is sufficiently high to assume the existence of an item-factor relationship (Gorsuch, 1983).
For the subscale dealing with accuracy in self-appraisal, data in the table indicate that none of the emic items performed as expected. That is, out of five items, none of them loaded high on its factor. The performance of the etic items, on the other hand, was better as reflected by the rate of items loading high in the first factor. However, two of the etic items defined the fifth dimension, even though in theory they tap the first dimension. For the sub scale "goal selection," in spite of high item loading, none of the etic and emic items performed as indicator of a second dimension.

For the remaining dimensions, both emic and etic items behaved as indicators of their respective dimensions. For example, all of the items dealing with problem solving behaviors loaded high in the second dimension. A similar pattern was observed for the others two dimensions. The third factor, gathering information, clustered nine out of the ten items. For this dimension, all of the newly developed items exhibited the most salient factor loading. Finally, dimension 4 measured skills necessary for making plans. With only one exception, the remaining items loaded high on it.

As part of the second stage of the study, the invariance across two populations of the factor structure was investigated with a sample of 400 high school students. Similar statistical analyses as those from the previous section were conducted for each group.

For the high certainty group, the factor structure accounted for 43.3% of the variability in the observed data. The root mean squared index was of size 0.049 indicating a good model fit. Small size of the root mean square is indicative of a small discrepancy between observed and adjusted data under the hypothesized model. The first two dimensions accounted for more than two-thirds of the observed data variability. Regarding the low certainty group, the adjusted model explained approximately two-fifths of the observed data variability. The root mean squared was 0.052. However, compared with previous results, approximately two-thirds of the low certainty student data variability was explained by the hypothesized model. Tables 5 and 6 summarize descriptive information of the sample and rotated factor pattern matrices for each of the two groups (certain vs. uncertain) used in this part of the study, respectively.

Insert Table 5 and Table 6 here

The analysis of factor congruency by means of procrustes transformation indicated noninvariance of the constructs. In these analyses, the structure devised in the first part of the study was chosen as

the target matrix and the source matrices were those from decided and undecided groups. Table 7 summarizes results for overall congruency and factor congruency.

Insert Table 7 here

From the above table, total structure congruency coefficients of about 0.89 and 0.87 for decided and undecided samples are indicative of congruency with the target matrix. According to Paunonen's rule of thumb, estimates larger than 0.42 are necessary to establish an overall factorial congruency between target and rotated structures. These results might indicate the possibility of matching the overall factorial structure of the career decision making self efficacy from the factor pattern matrix of both decided and undecided groups.

The congruence between target and rotated structure at the factor level was poor for the two rotated groups. For decided students two out of five comparisons were significantly different from random rotations. In Paunonen's rule of thumb, random solution can be ruled out by observing factor similarity greater than 0.90. Table 7 indicates that factor congruence was achieved only for the subscales dealing with information gathering and making plans. Another interesting finding is the poor congruency observed for the fifth factor, which is defined only by at most two salient items. For undecided students, factor congruency was even lower. Out of five dimensions, the one dealing with problem solving was significantly reproduced.

Finally, Table 8 and Table 9 summarize results for item congruencies for decided and undecided groups, respectively. Both tables group estimated item congruence coefficients using Paunonen's rule of thumb of 0.85 to support the contention of item invariance across groups. For decided students, congruence coefficients for about one fourth of the emic and etic items were higher than that of 95% of rotations from random data. Approximately fourth fifths of the emic items exhibited significant congruence coefficients. A similar proportion of etic items was found significant, also. Another important observation in Table 8 is that emic items with congruence coefficients higher than rotations from random data came from behaviors dealing with making plans, gathering information, and problem solving. Except for problem solving, the other two dimensions achieved congruence coefficients higher than that of 95% of rotations from random data.

For undecided students, congruence coefficients for about one half of the total number emic and etic items were significantly greater than those from random samples. Approximately three-fifths of the emic items and one half of the etic items showed significant congruence coefficients. Finally, five etic items and five emic items consistently were flagged with low congruence coefficients for decided and undecided groups.

CONCLUSION AND DISCUSSION

Adapting tests across countries involves more than adapting items linguistically and using judges to apprise items for cultural adequacy. It involves an assessment of the generalization of a construct theory. When transferring a psychological scale from one country to another, several hazards limit a direct interpretation of test scores. Some problems arise from particular decisions made during test construction (e.g., developing a test in a source language and then translating to a target language) and others from test development decisions (e.g., using item writers from the source culture).

The marginal effect of adding items to strive for culturally appropriateness was consistent with results from other studies (de Vera, 1985; Hulin, Drasgow, &Komoar, 1982). Tailoring a test from emic and etic items slightly improves reliability of scale scores. However, tailoring a scale from the best items available in the two cultures yielded some discouraging results. Contrary to hybrids from the natural world, those from psychometrics are not stronger than their "parents". Moreover, the tailored scale failed to recreate the structure for the construct to a point that construct and factor noninvariance was observed.

The inclusion of etic items in the scale required adapting item content to the target culture. By increasing generality of the behavior in the item, several items were more culturally relevant to Mexico. However, this strategy diminished the information gathered from the scale. For example, by phrasing the item "find information in the library about occupations you are interested in" as "find information about occupations you are interested in," cultural relevance was gained at the expense of information. A low score in the former item can be indicative of limited skills in using the library to seek occupational information as part of career decision making. This information could be used as part of the vocational orientation training process. However, a low score in the culturally adapted item might be indicative of several reasons, but none of them can be directly inferred.

On the other hand, developing good emic items is not as straightforward of a task. Even though a great amount of care is devoted to their construction, there is no guarantee that emic will perform as expected. In this research, emic items although developed as indicators for the goal selection dimension were consistently found defining a different dimension. On the other hand, for the gathering information subscale, emic items improved its quality.

There are other hazards coming from characteristics of target culture that often are overlooked for the test construction process, however. The extreme response style (ERS) is one of these possible situations. There are cultures characterized by being high in ERS, with their citizens wishing to demonstrate sincerity and conviction. Whereas, in cultures characterized by being low in ERS, citizens might want to be perceived as modest and nonjudgmental (Riordan and Vandenberg, 1994; Cheung and Rensvold, 2000). The behavior of the above two kinds of cultural patterns in a psychological scale in which respondents ought to scale themselves in a nine-point continuum would be different. The distribution of item response for the former group will be skewed to the right or to the left, whereas for the latter group will be near the center of the scale.

When extreme response style is present, the lack of equivalence of a construct in the two cultures might be confounded. Is the construct really noninvariant in culture A, is the extreme response

style operating to create such a lack of invariance, or is the combination of both? To disentangle the effects of both sources, two groups differing in their amounts of extreme response styles can be used, for example. In Mexico, it is not uncommon to expect those students who have made their decision to have a strong conviction to convey more sincerity and to socially impress others with this conviction. Unfortunately, the contention that extreme response style was causing the failure for reproducing the construct and factorial structure cannot be totally supported from findings in this study. For groups differing in their amounts of extreme response style, even though the one with the largest probability for exhibiting such kinds of responses matches partially the target structure. One possible might be connected to the samples involved in the study. The target structure was gathered from a sample that might have a mixture of decided and undecided students; thus, the partial match in structure might be due to its overlap with the other two samples. Other possible explanation might be the approach followed to gather samples of students exhibiting high and low amounts of extreme response style. A better approach could be using scores from a scale measuring extreme response style (Greenleaf, 1992). However, such a scale has not been translated into Spanish and there is no evidence that a score from it could be exempt of the effects of extreme response style.

Cross-cultural research involves several kinds of translation. Translation, the issue addressed in this article, is clearly presented as more than a linguistic problem, or even a problem of style. It is a deeper issue, which concerns the translation of intangible concepts and personal or collective experiences. This is where a problem arises, one, which is not only scientific, but one which is rooted on cultural knowledge and sensitivity. A similar situation consistently arises in other disciplines, such as anthropology. Neither educators nor anthropologists engaged in cross-cultural research have immediately at their disposal a "culture handbook" by which they might easily locate cultural differences and uniqueness.

REFERENCES

AERA, APA, NCME (1999). Standards for educational and psychological testing.

Allalouf, A., Hambleton, R., and Sireci, S. (1999). Identifying the causes of DIF in translated verbal items. Journal of Educational Measurement, 36, 185-198.

Angoff, W. and Cook, L. (1988). Equating the scores of the Prueba de Aptitud Academica and the Scholastic Aptitude Test (Research Report No. 3). New York: College Entrance Examination Board.

Arce-Ferrer, A. (2001). Factorial invariance in psychological testing: new construct dimensions or differential item interpretation? Revista Mexicana de Psicología, 18, 11.

Betz, N., Klein, K., and Taylor, K. (1996). Evaluation of a short form of the career decision-making self-efficacy scale. Journal of Career Assessment, 4, 47-57.

Betz, N., and Luzzo, D. (1996). Career assessment and the Career Decision-Making Self-Efficacy Scale. Journal of Career Assessment, 4, 413-428.

Brislin, R. (1970). Back-translation for cross-cultural research. Journal of Cross-Cultural Psychology, 1, 185-216.

Canto, J. (1999). Factors influencing certainty of career choice. Unpublished doctoral dissertation, Autonomous University of Yucatan, Mexico.

Chan, W., Ho, R., Leung, K., Chan, D., and Yung, Y. (1999). An alternative method for evaluating congruence coefficients with procrustes rotation: A bootstrap procedure. Psychological Methods, 4, 378-402.

Cheung, G. and Rensvold, R. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equation modeling. Journal of Cross.Cultural Psychology, 31, 187-212.

Clarke III, I. (2000). Extreme response style in cross-cultural research: An empirical investigation. Journal of Social Behavior and Personality, 15, 137-152.

Cliff, N. (1966). Orthogonal rotation to congruence. Psychometrika, 31, 33-42.

Crites, J. (1978). Career maturity inventory. Monterey, CA: CTB/McGraw-Hill.

Díaz-Guerrero, R. (1975). Psychology of the Mexican: Culture and personality. Austin: University of Texas Press.

DeVera, M. (1985). Establishing cultural relevance and measurement equivalence using emic and etic items. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

Fouad, N., Smith, P., and Enochs, L. (1997). Reliability and validity evidence for the middle school self-efficacy scale. Measurement and Evaluation in Counseling and Development, 30, 17-29.

Gibbons, J., Zeller, J., and Rudek, D. (1999). Effects of language and meaningfulness on the use of extreme response style by Spanish-English bilinguals. Cross-cultural Research, 33, 369-381.

Greenleaf, E. (1992). Measuring extreme response style. Public Opinion Quarterly, 56, 328-351.

Gorsuch, R. (1983). Factor analysis. Hillsdale, NJ: Erlbaum.

Hambleton, R. (1994). Guideliness for adapting educational and psychological tests: A progress report. European Journal of Psychological Assessment, 10, 229-224.

Hambleton, R. and Jones, R. (1995). Comparison of empirical and judgmental procedures for detecting differential item functioning. Educational Research Quarterly, 18, 21-36.

Hulin, C. (1987). A psychometric theory of evaluations of item and test translations: Fidelity across languages. Journal of Cross-Cultural Psychology, 67, 115-142.

Hulin, C., Drasgow, F., and Komoar, J. (1982). Applications of item response theory to analysis of attitude scale translations. Journal of Applied Psychology, 67, 818-825.

Kort, B. and Tucker, L. (1976). Procrustes matching by congruency coefficients. Psychometrika, 41, 361-372.

Little, T. (2000). On the comparability of constructs in cross-cultural research. A critique of Cheung and Rensvold. Journal of Cross.Cultural Psychology, 31, 213-219.

Luzzo, D. (1993). Reliability and validity testing of the Career Decision-Making Self-Efficacy Scale. Measurement and Evaluation in Counseling and Development, 26, 137-142.

McCrae, R. Zonderman, A., Costa, P., Bond, M., and Paunonen, S. (1996). Evaluating replicability of factors in the revised NEO Personality Inventory: confirmatory factor analysis versus procrustes rotation. Journal of Personality and Social Psychology, 70, 552-566.

Merenda, P. (1997). A guide to the proper use of factor analysis in the conduct and reporting of research: Pitfalls to avoid. Measurement and Evaluation in Counseling and Development, 30, 156-164.

Mullen, M. (1995). Diagnosing measurement equivalence in cross-national research. Journal of International Business Studies, 23, 169-179.

Paunonen, S. 1997). On chance and factor congruence following orthogonal procrustes rotation. Educational and Psychological Measurement, 57, 33-59.

Ramirez, M., Arce-Ferrer, A., and Canto, J. (2000). Students' Self-efficacy. Social Psychology in Mexico, 7, 297-302.

Rensvold, R. and Cheung, G. (1998). Testing measurement models for factorial invariance: A systematic approach. Educational and Psychological Measurement, 58, 1017-1034.

Riordan, C. and Vandenberg, R. (1994). A central question in cross-cultural research: Do employees of different cultures interpret work-related measures in a equivalent manner? Journal of Management, 20, 643-671.

Schönemann, P. (1966). A generalized solution to the orthogonal procrustes problem. Psychometrika, 31, 1-10.

Table 1:

Sample characteristics for the invariance of the self-efficacy construct

Age (years)	Certain about a major		Uncertain about a major	
	Male	Female	Male	Female
15	25	25	25	25
16	25	25	25	25
17	25	25	25	25
18	25	25	25	25
Total	100	100	100	100

Table 2:

Sources of items comprising the adapted version of the Career Decision-Making Self-Efficay Scale for Spanish speaking high school students from Mexico

Subscale	Etic	Emic	Total
Accurate self-appraisal	5	5	10
Gathering occupational information	5	5	10
Goal selection	5	5	10
Making plans for the future	5	6	11
Problem solving	5	4	9
Total	25	25	50

Table 3:

Values of the coefficient alpha for different versions of the Career Decision-Making Self-Efficay Scale

Subscale	Source Language		Target Language	
	CDMSE[a]	CDMSE-SF[b]	CDMSE-SF[c]	Emic and etic items[d]
Accurate self-appraisal	0.88 (10)[e]	0.73 (5)	0.60 (5)	0.79 (10)
Gathering occupational information	0.89 (10)	0.78 (5)	0.69 (5)	0.86 (10)
Goal selection	0.87 (10)	0.83 (5)	0.62 (5)	0.87 (10)
Making plans for the future	0.89 (10)	0.81 (5)	0.73 (5)	0.88 (10)
Problem solving	0.86 (10)	0.75 (5)	0.55 (5)	0.87 (10)
Total	0.97 (50)	0.94 (25)	0.88 (25)	0.95 (50)

(a) Career Decision-Making Self-Efficay Scale. Reliability estimates came from the test manual
(b) Career Decision-Making Self-Efficay short form scale. Reliability estimates came from the test manual
(c) Spanish version of the Career Decision-Making Self-Efficay short form scale
(d) Emic and etic items comprising the tailored version of the Career Decision-Making Self-Efficay scale
(e) Number of items

BEST COPY AVAILABLE

Table 4:

Varimax rotated factors from the adapted version Career Decision-Making Self-Efficay Scale

		Items in each sub scale									
Subscale		1	2	3	4	5	6	7	8	9	10
Accurate self-appraisal	F1 F5	.40	.58	a,b	.42	a,b	a,b	a,b	a,b	b .64	b .79
Goal selection	F1	.66	.65	.62	.64	.52[a]	.54[a]	.50[a]	.45[a]	.42[a]	b
Problem solving	F2	.58	.47	.44	.55	.59[a]	.69[a]	.62[a]	.57[a]	.59	.63
Gathering information	F3	.46	.55	.42	.74[a]	.58[a]	.71[a]	.49[a]	.54[a]	b	.44
Making plans	F4	.48	.49	.58[a]	.61[a]	.65[a]	.53	.53[a]	.55[a]	.42[a]	b

(a) Emic items
(b) Factor loadings smaller than 0.40 in absolute value

Table 5:

Varimax rotated factors from the adapted version Career Decision-Making Self-Efficay Scale for sample high school students with a chosen major

		Items in each sub scale									
Subscale		1	2	3	4	5	6	7	8	9	10
Accurate self-appraisal	F1	[b]	.56	[a,b]	[b]	.54[a,b]	.77[a,b]	[a,b]	[a,b]	[b]	.44
	F5									.41	[b]
Goal selection	F1	.43	.46	.49	.46	.62[a]	.69[a]	.54[a]	.45[a]	.43[a,b]	[b]
Problem solving	F2	.82	.41	.44	[b]	.46[a]	.74[a]	.86[a]	.41[a]	[b]	[b]
Gathering information	F3	.57	.47	.50	.45[a]	.45[a]	.42[a]	.47[a]	.51[a]	.41	.61
Making plans	F4	.58	.44	.65[a]	.57[a]	.46[a]	.43	.45[a]	[a,b]	[a,b]	[b]

(a) Emic items
(b) Factor loadings smaller than 0.40 in absolute value

Table 6:

Varimax rotated factors from the adapted version Career Decision-Making Self-Efficay Scale for sample high school students with unchosen major

		Items in each subscale									
Subscale		1	2	3	4	5	6	7	8	9	10
Accurate self-appraisal	F1	[b]	[b]	.55[a]	[b]	[a,b]	[a,b]	[a,b]	[a,b]	[b]	[b]
	F5									[b]	[b]
Goal selection	F1	.53	.64	.61	.49	.57[a]	[a,b]	.56[a]	.55[a]	.49[a,b]	.63
Problem solving	F2	[b]	.68	[b]	[b]	.49[a]	.75[a]	[a,b]	.51[a]	.57	.41
Gathering information	F3	.85	.45	[b]	.46[a]	.51[a]	.72[a]	.87[a]	.41[a]	.49	.76
Making plans	F4	.66	[b]	.45[a]	[a,b]	[a,b]	[a,b]	.56[a]	[a,b]	[a,b]	[b]

(a) Emic items
(b) Factor loading smaller than 0.40 in absolute value

Table 7:

Total and factor congruence coefficients for samples of decided students and undecided high school students with respect to a sample with both kinds of students

Sample	F1	F2	F3	F4	F5	Total
Decided students	0.88	0.88	0.91*	0.95*	0.80	0.89*
Undecided students	0.88	0.90*	0.87	0.84	0.86	0.87*

* Congruence higher than that of 95% of rotations from random data
F1 = Accurate self-appraisal
F2 = Problem solving
F3 = Information gathering
F4 = Making plans
F5 = Goal selection

Table 8:

Variable congruence coefficients for sample of decided school students with respect to a sample with both decided and undecided students

Status of the Congruence	Self-appraisal Goal selection	Problem Solving	Gathering information	Making plans
Higher than that of 95% of rotations from random data	2,**5**,**7**,9,10,11,13,14,**15**,**16**,**17**,**18**,20	21,22,23,**25**,**26**,**27**,**28**,29	31,33,**34**,**35**,**36**,**37**,**38**	41,42,**43**,**44**,**45**,**47**,**49**,50
Lower than that of 95% of rotations from random data	1,**3**,4,**6**,**8**,12,**19**,	24,30	32,39,40	46,**48**

Emic items are in bold face.

Table 9:

Variable congruence coefficients for sample of undecided students with respect to a sample with both decided and undecided high school students

Status of the Congruence	Self-appraisal Goal selection	Problem Solving	Gathering information	Making plans
Higher than that of 95% of rotations from random data	1,**5**,**7**,9,10,11,12,13,14,**15**,**16**,**17**,**18**	22,**26**,**28**,29,30	31,**34**,**35**,**36**,**37**,**38**	41,**43**,**47**
Lower than that of 95% of rotations from random data	2,**3**,4,**6**,**8**,**19**,20	21,23,24,**25**,**27**,	32,33,39,40	42,**44**,**45**,46,**48**,**49**,50

Emic items are in bold face.